

# Steps to set up the Data Representation WG and transition from the FITS WG

Draft 0.2 prepared by L.Chiappetti – IASF Milano – 09 Sep 2016

## Background

The IAU FITS Working Group (hereafter FITS WG or FWG) was created in 1988 by Resolution B2 at the XXth IAU General Assembly at Baltimore *to maintain the existing FITS standards and to review, approve and maintain future extensions to FITS, recommended practices for FITS implementations, and the thesaurus of approve FITS keywords*. It served under Commission 5 (Documentation and Astronomical Data) of IAU Division B.

With the re-arrangement of IAU Commissions which took place at the XXIXth IAU General Assembly at Honolulu in August 2015, the tasks of the former commission are now assigned to the new Commission B2 Data and Documentation, which articulates in the following Working Groups:

- WG1. Designations
- WG2. Virtual Astronomy and Data Centers
- WG3. Data Preservation and Curation
- WG4. Data Representations**
- WG5. Software Practices, Verification and Validation
- WG6. Data Based Astronomy Education and Public Outreach

With some surprise I found to have been nominated at the chair of WG4, Data Representation WG (hereafter DRWG), which, not surprisingly, should succeed to the FWG according to the proposal for continuation of former Commission 5 submitted at due time by Bob Hanisch (former President of Commission 5).

The operation of IAU Commissions and their establishment of Working Groups is described at the following link [http://www.iau.org/administration/statutes\\_rules/working\\_rules/#WorkRulX](http://www.iau.org/administration/statutes_rules/working_rules/#WorkRulX) (in particular see points 40 and XII, working group members are usually IAU members, but individuals who are not IAU members can be *nominated to the status of Associates by the Working Group and approved by the relevant Commission President*, point 47a).

I note that Working Groups are usually fixed term structures which cease to exist and/or have to be recreated at each General Assembly. However the FWG continued to operate since 1988 and was recognized an obvious candidate as a functional WG in reply to the IAU Questionnaire in Jan 2013. The actual status of functional WGs (permanent, long term) within IAU is unclear to me.

## Terms of reference for the DRWG

My only reference is the proposal submitted by Bob Hanisch on 31 Jan 2015, which reads:

*2.5 A WG for Data Representations (an expanded WG FITS) will help ensure and maintain the fluent interoperability of telescope data that has not only made multi-wavelength astronomical research commonplace, but has also made astronomy's data management practices the envy of many other disciplines. The WG FITS has been the custodian and advocate for the Flexible Image Transport System format since it was formally endorsed by the IAU in 1982; however, the more recent data landscape has broadened substantially, and many new facilities are exploring alternatives to the FITS standard in order to manage their issues of data scale and complexity. The WG FITS will therefore be expanded, and renamed Data Representations. It is vital to manage a careful and minimally disruptive transition from FITS to more modern and capable data representations, and in order to assure that continuity we plan to retain the FITS governance structure through a FITS Subcommittee*

*that is part of the new WG. Subcommittees focused on other data representation standards can also be created as needed.*

### *3. Associates*

*The remit of this Commission encompasses the understanding of astronomical data in many forms, and requires both scientific and technical expertise. Commission 5 routinely relied upon Associates, particularly in the WG Libraries and WG FITS, and we expect that need to continue.*

## **Models of Working Group operations**

There are essentially two models of operation for a working group: one is a broad forum for discussions with a large number of members (a good example could be the Time Domain Astronomy WG, which is a Division B WG and has 128 members); the other is a restricted group of members with a specific task and rules. The latter model was followed so far by the FWG (which at end-of-life had 22 members, and very specific voting rules).

However the FWG exploited, as a liaison to the astronomical community (with no restriction about IAU membership) a mailing list (originally gatewayed to an Usenet newsgroup) [fitsbits@nrao.edu](mailto:fitsbits@nrao.edu). Previously to 2014 the FWG had also a structure of Regional Committees (see the FWG web site for historical details).

## **A mixed model for the DRWG**

I cannot think for the future DRWG to any other model than a two-tier one which mixes the above two in a scheme like the following:

- a) an **outer tier** composed by interested individual members (should they be normally IAU members ? Can any member apply without screening ? Can a non member apply for participation ?) which is essentially a discussion forum (not unlike fitsbits, but with a broader scope)
- b) a **number of** sub-committees (here after **SEGs, Special Expert Groups**, according to a proposal by Preben Grosbol) with restricted membership and rules, each one in charge of a specific data format or other topic.

The FWG will naturally evolve in a FITS SEG, which may start acting immediately, and would *represent a prototype* for future other SEGs (e.g. for the rules).

While the establishment of the FITS SEG would represent a natural way to implement some long awaited or due changes in membership and rules (see next section), it is not clear to me how to proceed for the establishment of the outer tier and the proposal of new SEGs.

*For populating the outer tier*, I think we should issue a poll/invitation within IAU members (do commission and division levels mailing lists exist and are they open for such use ?) and/or at other national or interest community level (announcing on existing lists or groups). I will call for suggestions by other FWG members and by the members of the Commission B2 OC.

*For creation of new SEGs* other than the FITS SEG we should most collect suggestions from the members (or indications from Commission B2). Some examples are presented in the next subsection.

Now talking about the organization of the DRWG

I do not feel particularly confident in myself covering the task of DRWG chair on a long term basis.

I could do that ad interim until Commission B2 nominates a definitive chair, also depending on the evolution of the SEGs. I would instead be happy to continue chair the FITS SEG at least until the end of the current term (2018).

A procedure shall be defined to nominate the chair (and vice-chair if wished) of the DRWG (and the SEG membership, see below). So far a co-optation mechanism was used de facto.

Most likely in the regime phase the DRWG will need a small Executive Committee. A natural choice would be to form it from the SEG chairs (but it won't be obvious until we have a few different SEGs) and vice-chairs if any. To be decided if the DRWG chair can chair also one of the SEGs or if we want incompatibility.

It has to be decided whether the DRWG (or its Executive) shall have ex-officio members drawn from other Commission B2 WGs (I am thinking of WG2 and WG5) as liaison officers, mimicking a practice used in the FWG (see below).

Concerning operational tools of the DRWG, I foresee at least a WG web site and a WG mailing list. While I am technically capable of managing such tools (not the same could be said of different tools if proposed), please note what follows: (a) there are no technical objections for me to maintain an apache web site but I do not plan to have the time to do it; (b) while I am capable of setting up and maintaining a mailman mailing list on my machine (actually I do so for some FWG internal task forces) there is a little technical issue due to local arrangements (they have to be of the form [list@machine.domain](mailto:list@machine.domain) and cannot be of the form [list@domain](mailto:list@domain)), so I would prefer if *these tasks are delegated to some other hosting institution*. Or should we host them on a IAU site ? Or use some other way (e.g. a wiki on pbworks ?)

I assume the tools of the FWG (a web site hosted and managed at HEASARC; a main mailing list hosted at NRAO and managed by me; plus occasionally ad-hoc lists on my site or at HEASARC) will continue to operate as now.

## ***A list of possible SEGs***

The input to this section was provided mainly by Dick Shaw (verbatim in italics). More inputs welcome.

### **SEG 0: FITS SEG**

Purpose: Handle the curation of FITS format succeeding the FWG, as described further below.

### **SEG n: xxxx data format SEG**

Purpose: Handle the curation of each other individual data format xxxx which may be widely used by the astronomical community or a significant subset thereof. If any. Suggestions welcome.

### **SEG 1 : Next generation data format SEG**

*Purpose: design and prototype a new data format that addresses current and future data format needs, leveraging the successes and avoiding the limitations of FITS format. The end goal would be to formulate a proposal for a next-generation standard.*

### **SEG 2 : Data provenance SEG**

*Purpose: Construct a mechanism for establishing and maintaining data provenance from product creation through publication, replication, and archival storage. Some of this work is already going on, but as far as we know there is no IAU body that has this topic as an area of concern.*

### **SEG 3 : Event stream SEG**

*Purpose: Formalize existing event protocols, archiving, semantics, etc. This is an obvious connection to the VO work in this area, but with the goal of transitioning from research projects to an IAU standard.*

### **Organization of the SEGs**

*Some SEGs have an obvious, near term purpose and should therefore have a limited life-span. Their charter and duration would be specified in advance (with possible extensions by the DRWG). Other SEGs might have a longer-term, curation role.*

*The membership shall be limited in number, but with criteria depending on the SEG.*

*For long term curation SEGs the membership (rather than the SEG) could have a finite term, like two IAU triennial mandates, possibly with a fraction of the group being replaced each year. In this case the time commitment would typically be rather small (except perhaps for occasional task forces).*

*For a short term SEG with a specific task, like the Next Generation Data Format SEG, one would need members with a high level of technical and domain knowledge, and who actually have significant time to devote to the effort during the finite life of the SEG. The group would need a chair to keep the effort organized, and membership should probably not exceed a dozen persons. Progress would need to be visible to the community. Community members should be able to contribute ideas, but should not expect to be involved in the decision making (this is not unlike the model followed by the FWG with FITSBITS and the FWG itself). Such SEG's end result is a well developed proposal, which would be subject to community review and discussion via the DRWG through some TBD mechanism.*

*The voting rules could be different, depending upon the charter of the SEG, but we give below a possible model based on the FWG.*

## **Proposal for an updated FWG as model SEG**

### **Specific FWG issues**

The final composition of the FWG consisted of 22 members, appointed (co-opted ?) in an epoch variable from 1988 to 2010 (see [http://fits.gsfc.nasa.gov/iaufwg/iaufwg\\_members.html](http://fits.gsfc.nasa.gov/iaufwg/iaufwg_members.html)), but this in the process of being re-arranged autonomously and the relevant information has been removed from this draft.

The retirement or emeritus status at the home institution has not forbidden so far anybody from being an active (voting) FWG member.

The `iaufwg` mailing list has traditionally included as observers (non-voting members) some past members (mostly retired), a “Designated FITS Representative” from the former VO WG (now succeeded by WG2), and the President of the overarching commission (5). De facto some of these roles often coincided.

The FWG had an Executive Committee (EC) of 7 members, which was traditionally based on the (now disbanded) Regional Committees. De facto the need for the EC to act separately from the plenary WG is virtually disappeared.

Also the FWG had for most of its lifetime a vice-chair designed to succeed to the chair in a future term. This has been discontinued recently because of the IAU rearrangement.

Concerning these general matters I suggest that the FITS SEG (or any other SEG) has *no longer an*

*executive committee*, while the presence of a *vice-chair should be re-instated* (this could be a model for long term curation SEGs). I also propose that the iaufwg and fitsbits mailing lists continues to serve for the FITS SEG, with the current practice (i.e. allowing significant past members as observers in the iaufwg list, and usage of fitsbits as forum for the community beyond IAU). It is recommended that the chair of the DRWG (if not a SEG member) be an observer. It is up to the President of Commission B2 to decide whether he wants to be an observer on the FITS SEG mailing list or just on the future DRWG list.

## **The ideal SEG**

What follows is a free re-elaboration of material provided by Preben Grosbol and Adam Dobrzycki (in italics when copied verbatim mainly from a draft tagged 1.1 2015-03-09)

*The SEGs should have a controlled membership to ensure that 'important' organizations and experts are represented in a fair way.*

*The DRWG or its Executive should establish voting rules for the recommendation of standards. They should include consent from both the general discussion forum and the relevant SEG. A modified set of the FWG voting rules could be used as a guide*

The FITS SEG shall be composed by 20-25 members. This is just a guideline for other SEGs but has proven effective during the FWG lifetime. Specific short term “intensive” SEGs could have a membership of half that number, about a dozen.

*The current voting rules for the FWG were designed to ensure that a new FITS format would be endorsed by a significant majority (i.e. 3/4) and that any voice of serious concern would be dealt with by re-discussing the proposal. [...] A member of the FWG may either represent an organization or an individual even though all votes have equal weight. This [could, although it never happened,] lead to the rejection of a proposal if a few single persons either vote against or refrain from voting even if all major organizations would vote in favor.*

The idea is that a SEG (starting with the FITS SEG) be composed by institutional members and individual members. *All members of the SEG have a personal vote*, but institutional members' vote weights differently. The approval of a proposal for change of the FITS standard will require yes-votes *by a simple majority (50%) of all registered votes (i.e. personal and institutional votes cast), and by 3/4 of the institutional votes. If there are institutional no-votes during the first vote for a proposal, the proposal cannot be adopted before it is re-discussed. A second vote cannot occur before 3 months after the first vote.*

This places a burden on the institutional members. They are *expected to read, comment, and vote on proposals brought up for voting. If such persons do not cast a vote, the chair of the SEG can request to have the person replaced. An organization can request to have their organizational vote transferred to another person by request to the chair of the SEG.*

(in the above sentences I replaced the SEG Executive with the chair of the SEG since I do not foresee an executive; otherwise we may think of an Executive composed by the chair, vice-chair and institutional members. I wonder whether we need to put incompatibility rules, like the fact the chairperson shall be drawn from the individual members and not be an institutional member).

Concerning the membership of a SEG should we formalize how (individual) members are selected, or just proceed by co-optation as done so far ?. Can anybody (IAU member) just opt-in, or shall one be proposed, approved or invited (experts at Associate level should be invited) ? And by whom (SEG chair, SEG members, DRWG chair, Commission B2 ?).

Concerning institutional members there are several issues:

What is the list of “major data handling organizations” entitled to an institutional member ?

Do we need some formal document from an authority within the organization which nominates the institutional representative ?

And in general how formal shall we be (for instance if there are two members affiliated to an institution can they transfer the institutional vote among them in case one is not available ? Can an institution be represented by an emeritus member ?)

How do you define a major organization ? Is it an observatory or data centre managing data (\*) above a given amount ? Or should also we consider teams in charge of large software packages (\*) ?

(\*) data (and packages dealing with it) in the specific format covered by the SEG

Do we need additional considerations of “regional” or national balances ?

A first definition of the institutional members may start from the current membership. More institutions can be suggested (by whom ?) and should receive *support from a 3/4 majority of the current institutional votes*.

Looking at the current membership, two space agencies (NASA and ESA) are covered, as well as three major observatories (ESO, NOAO and NRAO) and a major data centre (CDS), or more if others feel major ...as I noted STScI is currently not represented. We have national representations of the USA (abundant), Canada, UK, Netherlands, France, Italy, Russia, Japan and Australia. We have I feel a fair coverage of disciplinary fields and wavebands. Concerning (non mission-specific) software packages the following are represented (it is not always obvious when some members represent a software package, an institution or both): AIPS, IRAF, CFITSIO, WCSlib.

Should we aim to a coverage at “mission” level (e.g. space satellite or large projects like SKA or CTA) when defining major institutions ? What is missing at software level (topcat ?) Should we enlarge national/regional representations (China, India, South America, Germany, Spain ?)

## Action plan

I circulated a first draft of this document inside to the FWG, and to the President and OC of Commission B2. I also proposed a BoF session "FITS (Flexible Image Transport System) and Data Representations WG" for ADASS XXVI in Trieste (October 2016) and this has been accepted and will take place for about 90 min after one of the afternoon plenary sessions.

This draft 0.2 is now circulated to the FITSBITS exploder, and the recipients are encouraged to spread the news in other forums. The initial discussion can take place on FITSBITS, then face to face at ADASS, awaiting the establishment of a proper online forum.

In the meanwhile the rearrangement of the FITS SEG membership will occur autonomously within the current FWG framework, as first operating nucleus of the DRWG.

## Document history

Draft 0.0	09 Jun 2016	original, circulated to IAUFWG
Draft 0.1	13 Jun 2016	minor revision (typos), circulated to Commission B2 OC
Draft 0.2	09 Sep 2016	simplified, removed FITS specifics, circulate to FITSBITS